Stationary Markov chains with linear regressions

Wlodzimierz Bryc
Department of Mathematics
University of Cincinnati
PO Box 210025
Cincinnati, OH 45221–0025
Wlodzimierz.Bryc@UC.edu *

February 1, 2008

Abstract

In Bryc(1998) we determined one dimensional distributions of a stationary field with linear regressions (1) and quadratic conditional variances (2) under a linear constraint (7) on the coefficients of the quadratic expression (3). In this paper we show that for stationary Markov chains with linear regressions and quadratic conditional variances the coefficients of the quadratic expression are indeed tied by a linear constraint which can take only one of the two alternative forms (7), or (8).

1 Introduction

Let $(X_k)_{k\in\mathbb{Z}}$ be a square-integrable random sequence. Consider the following two conditions.

$$E(X_k|\ldots,X_{k-2},X_{k-1},X_{k+1},X_{k+2},\ldots)=L(X_{k-1},X_{k+1}) \tag{1}$$

for all $k\in\mathbb{Z}$.

$$E(X_k^2|\ldots,X_{k-2},X_{k-1},X_{k+1},X_{k+2},\ldots)=Q(X_{k-1},X_{k+1}) \tag{2}$$

for all $k\in\mathbb{Z}$.

A number of papers analyzed conditions similar to (1) and (2). Of particular interest are papers Wesolowski(1989) and Wesolowski(1993), who analyzed continuous time processes X_t with linear regressions and quadratic second order conditional moments $Q()$ under the assumption that variances of X_t are strictly increasing; these processes turned out to have independent increments. Szablowski(1989) relates distributions of mean-square differentiable processes to conditional variances. Bryc & Plucinska(1983) show that linear regressions and constant conditional variances characterize gaussian sequences. In Bryc(1998) we show that a certain class of quadratic functions Q determines the univariate distributions for stationary processes which satisfy (1) and (2) with linear L. For additional references the reader is referred to Bryc(1995).

In this paper we assume that (X_k) is strictly stationary and the regressions are given by a symmetric linear polynomial $L(x,y)=a(x+y)+b$, and a general symmetric quadratic polynomial

$$Q(x,y)=A(x^2+y^2)+Bxy+C+D(x+y) \tag{3}$$

The linear polynomial $L()$ is determined uniquely by the covariances of (X_k). Namely, if the random variables X_k are centered with variance 1, the correlation coefficients $r_k=corr(X_0,X_k)$, and $r_2>-1$, then $L(x,y)=\frac{r_1}{1+r_2}(x+y)$. Since the moments of both sides of (2) must match, after standardization we also get the trivial relation

$$C=1-2A-Br_2 \tag{4}$$

*

Key Words: conditional moments, polynomial regression, linear regression
AMS (1991) Subject Classification: 60E99

This still leaves three parameters A, B, and D undetermined.

In this paper we analyze in more detail which quadratic polynomials $Q()$ can occur in (2) when (X_k) is a stationary Markov chain. We show that in this case we necessarily have $D = 0$ and that the remaining two coefficients satisfy one of the two linear equations (7) or (8). We show that if condition (7) is satisfied then the remaining free coefficient satisfies certain inequalities; under additional assumption (1), (2), and (7) characterize certain Markov chains uniquely.

2 Results

Through the rest of the paper we assume that (X_k) is standardized, $E(X_k) = 0, E(X_k^2) = 1$. We denote the correlations by $r_k := E(X_0X_k)$, $r := r_1$.

For Markov chains the regression equations (1) and (2) become respectively

$$E(X_k|X_{k-1}, X_{k+1}) = L(X_{k-1}, X_{k+1}) \tag{5}$$

$$E(X_k^2|X_{k-1}, X_{k+1}) = Q(X_{k-1}, X_{k+1}) \tag{6}$$

The following result shows that the coefficients of (3) are tied by a linear constraint.

Theorem 2.1 *Let (X_k) be a square-integrable standardized stationary homogeneous Markov chain such that $r \neq 0$, and $2|r| < 1 + r_2$. If (X_k) satisfies conditions (5) and (6), then the coefficients of $Q()$ in (3) satisfy $D = 0$ and either*

$$A(r^2 + 1/r^2) + B = 1 \tag{7}$$

or

$$2A + Br^2 = 1 \tag{8}$$

(When Q is non-unique this should be interpreted that there is a quadratic function Q with the coefficients satisfying $D = 0$ and at least one of the identities (7) or (8).)

It turns out that (7) implies additional restrictions on the range of the remaining free parameter A.

Theorem 2.2 *Let (X_k) be a standardized strictly stationary square-integrable sequence such that conditions (1) and (2) hold true, and the correlation coefficients satisfy $r \neq 0$, and $2|r| < 1 + r_2$. Suppose that the coefficients of quadratic form $Q()$ in (3) are such that $D = 0$ and (7) holds true.*

Then either $A \geq 1/(1 + r^2)$ or $A \leq \frac{r^2}{1+r^4}$.

The next theorem is a version of Bryc(1998), Theorem 2.1.

Theorem 2.3 *Suppose that (X_k) satisfies the assumptions of Theorem 2.2, and $\frac{r^2}{(1+r^2)^2} \leq A \leq \frac{r^2}{1+r^4}$. Then X_k is a Markov chain with uniquely determined distribution.*

One can also show that condition (8) implies that $|X_k| = |X_0|$ with probability one.

3 Two-valued Markov chains

Verification of condition (5) for two-valued Markov chains is a simple exercise. We include it here because two-valued chains play a role in the proofs of Theorem 2.1 and Proposition 4.1. They also occur as "degenerate cases" in linear regression problems: in Bryc(1998) we construct Markov chains that satisfy (5) and (6) for $A < r^2/(1 + r^4)$; the boundary value $A = r^2/(1 + r^4)$ corresponds to the two-valued case.

We consider only standardized chains with mean 0 and variance 1. Under this assumption, if a transition matrix is defined by

$$\Pr(a, a) = 1 - \alpha, \Pr(a, b) = \alpha, \Pr(b, a) = \beta, \Pr(b, b) = 1 - \beta \tag{9}$$

then the invariant distribution assigns probabilities

$$\mu(a) = \frac{\beta}{\alpha+\beta},\ \mu(b) = \frac{\alpha}{\alpha+\beta} \tag{10}$$

and the two values of the chain are

$$a = \sqrt{\frac{\alpha}{\beta}}, b = -\sqrt{\frac{\beta}{\alpha}} \tag{11}$$

We consider non-degenerate Markov chains with the correlation coefficient $r \neq 0, \pm 1$ only. This excludes three uninteresting cases: i.i.d sequences, constant sequences with $X_k = X_0$ for all k, and alternating sequences with $X_k = (-1)^k X_0$ for all k.

Proposition 3.1 *If (X_k) is a two-valued stationary Markov chain with the one-step correlation coefficient $r \neq 0, \pm 1$ then (X_k) satisfies condition (5) if and only if X_0 is symmetric with values ± 1.*

Proof. First notice that $\alpha\beta > 0$, so the values and probabilities in (10) and (11) are well defined. Indeed, if $\alpha\beta = 0$ then we have $X_k = X_{k-1}$ and hence $r = 1$.

A simple computation using (9-11) shows that the one-step correlation coefficient is $r = 1 - \alpha - \beta$, and the two step correlation is $r_2 = r^2$. Since by assumption $0 < |r| < 1$, this implies that $\alpha + \beta < 2$ and $\alpha + \beta \neq 1$.

By routine computation we get the following conditional probabilities

$$\begin{aligned}
\Pr(X_k = a | X_{k-1} = a, X_{k+1} = b) &= \frac{1-\alpha}{2-\alpha-\beta} \\
\Pr(X_k = b | X_{k-1} = a, X_{k+1} = b) &= \frac{1-\beta}{2-\alpha-\beta}
\end{aligned}$$

Using (5) we have $E(X_1|X_0 = a, X_2 = b) = \frac{r}{1+r^2}(X_0 + X_2) = \frac{\alpha-\beta}{\sqrt{\alpha\beta}} \frac{1-\alpha-\beta}{1+(1-\alpha-\beta)^2}$. On the other hand, direct computation using conditional probabilities gives $E(X_1|X_0 = a, X_2 = b) = \frac{\alpha-\beta}{\sqrt{\alpha\beta}} \frac{1-\alpha-\beta}{2-\alpha-\beta}$. The resulting equation has four roots when solved for β: the double root $\beta = 1 - \alpha$ and two roots $\beta = \pm\alpha$. Solution $\beta = 1 - \alpha$ corresponds to the independent sequence with $r = 0$. Since $\beta \geq 0$, therefore the only non-trivial solution is $\beta = \alpha$, which gives $p = \frac{1}{2}$ and $X_k = \pm 1$.

Condition (5) in this case is verified by direct computation with conditional probabilities. $\square$

4 Auxiliary results and proofs

Condition (1) determines the form of the covariance matrix $r_k = E(X_0 X_k)$.

Lemma 4.1 *Suppose that (X_k) is an L_2-stationary sequence such that condition (1) holds true and $2|r| < 1 + r_2$. Then $corr(X_0, X_k) = r^k$.*

Proof. Indeed, multiplying (1) by X_0 we get $r_k = a(r_{k-1} + r_{k+1})$. In particular, if $r := r_1 = 0$ then $a = 0$ and $r_k = 0$ for all k. On the other hand, if $r \neq 0$, then $1 + r_2 > 0, a = r_1/(1 + r_2)$ and the correlation coefficients r_k satisfy the recurrence

$$(1 + r_2) r_k = r(r_{k-1} + r_{k+1}), k = 1, 2, \ldots$$

From this we infer that $r_k \to 0$ as $k \to \infty$. Indeed, since $|r_k| \leq 1$, $r_\infty = \limsup_{k\to\infty} |r_k|$ is finite, and satisfies $r_\infty(r_2 + 1) \leq 2r_\infty|r|$. Is is easy to see that since $r_k \to 0$, the recurrence has unique solution $r_k = r^k$. $\square$

We use the notation $E(\cdot|\ldots, X_0)$ to denote the conditional expectation with respect to the sigma field generated by $\{X_k : k \leq 0\}$.

The following Lemma comes from Bryc(1998); the proof is included for completeness.

Lemma 4.2 *If (X_k) satisfies the assumptions of Lemma 4.1, then*

$$E(X_1|\ldots,X_0) = rX_0 \tag{12}$$

Proof. By Lemma 4.1, we have $r_k = r^k$, and $|r| < \frac{1+r^2}{2} \le 1$.

We first show by induction that for all $n \in \mathbb{Z}, k \in \mathbb{N}, 0 \le i \le k$

$$E(X_{n+i}|\ldots,X_{n-1},X_n,X_{n+k},X_{n+k+1},\ldots) = a(i,k)X_n + b(i,k)X_{n+k} \tag{13}$$

where $a(i,k) = \frac{r_i - r_{k-i}r_k}{1-r_k^2}$, $b(i,k) = \frac{r_{k-i}-r_ir_k}{1-r_k^2}$

For $k=2$, (13) follows from (1) when $i=1$. Clearly, (13) trivially holds true when $i=0$ or $i=k$ for all k.

Suppose that (13) holds true for a given value of $k \ge 2$ and all $n \in \mathbb{Z}$. We will prove that it holds true for $k+1$. We only need to show that the left-hand side of (13) is a linear function of the appropriate variables. Indeed, in the non-degenerate case the coefficients $a(i,k), b(i,k)$ in a linear regression are uniquely determined from the covariances; the covariance matrices are non-degenerate since $|r|<1$ and $r_k = r^k$.

Using routine properties of conditional expectations, the case of general index $0 < i < k$ reduces to two values $i = 1, k-1$. By symmetry, it suffices to give the proof when $i=1$.

Conditioning on additional variable X_{n+k} we get

$$E(X_{n+1}|\ldots,X_{n-1},X_n,X_{n+k+1},X_{n+k+2},\ldots) =$$
$$E(E^{\ldots X_{n-1},X_n,X_{n+k},X_{n+k+1},\ldots}(X_{n+1})|\ldots X_{n-1},X_n,X_{n+k+1},\ldots) =$$
$$a(1,k)X_n + b(1,k)E(X_{n+k}|\ldots,X_{n-1},X_n,X_{n+k+1},\ldots)$$

Now adding X_{n+1} to the condition we get.

$$E(X_{n+k}|\ldots,X_{n-1},X_n,X_{n+k+1},X_{n+k+2},\ldots) =$$
$$E(E^{\ldots,X_n,X_{n+1},X_{n+k+1},X_{n+k+2},\ldots}(X_{n+k})|\ldots,X_{n-1},X_n,X_{n+k+1},X_{n+k+2},\ldots) =$$
$$a(k-1,k)E(X_{n+1}|\ldots,X_{n-1},X_n,X_{n+k+1},X_{n+k+2},\ldots) + b(k,k)X_{n+k+1}$$

This gives the system of two linear equations for $E(X_{n+1}|\ldots X_{n-1},X_n,X_{n+k+1},X_{n+k+2},\ldots)$, which has the unique solution which is a linear function of X_n, X_{n+k+1} when $a(k-1,k)b(1,k) \neq 1$. It remains to notice that if $k>1$ then $a(k-1,k) = b(1,k) = \frac{r^{k-1}-r^{k+1}}{1-r^{2k}} < 1$. Indeed, the latter is equivalent to $r^{k-1}(1-r+r^{k+1}) < 1$ and holds true because $-1 < r < 1$, $r^{k+1} < 1-r+r^{k+1}$, and $1-r+r^{k+1} \le 1-r+r^2 \le 1$.

Therefore the regression $E(X_{n+1}|\ldots X_{n-1},X_n,X_{n+k+1},X_{n+k+2},\ldots)$ is linear, and (13) holds for $k+1$. This proves (13) by induction.

Passing to the limit as $k \to \infty$ in (13) with $n=0, i=1$ we get (12). $\square$

The following result comes from Bryc(1998). Since certain minor details differ we include it here for completeness.

Lemma 4.3 *If (X_k) satisfies the assumptions of Lemma 4.1 and (2) holds true, then*

$$(1 - A(1+r^2))E(X_1^2|\ldots,X_0) = (A(1-r^2)+Br^2)X_0^2 + C + D(1+r^2)X_0 \tag{14}$$

Proof. By Lemma 4.1 we have $L(x,y) = \frac{r}{1+r^2}(x+y)$. Since $E(X_1X_2|\ldots,X_0) = E^{\ldots,X_0}\left(X_1E^{\ldots,X_1}(X_2)\right)$, from Lemma 4.2 we get

$$E(X_1X_2|\ldots,X_0) = rE(X_1^2|\ldots,X_0) \tag{15}$$

We now give another expression for the left hand side of (15). Substituting $E(X_1X_2|\ldots,X_0) = E(X_2E(X_1|\ldots,X_0,X_2,\ldots)|\ldots,X_0) =$ into (5) we get $E(X_1X_2|\ldots,X_0) = \frac{r}{1+r^2}E(X_2(X_2+X_0)|\ldots,X_0)$.

By Lemma 4.2 this implies $E(X_1X_2|\ldots,X_0) = \frac{r^3}{1+r^2}X_0^2 + \frac{r}{1+r^2}E(X_2^2|\ldots,X_0)$. Since $r \neq 0$, combining the latter with (15) we have

$$E(X_2^2|\ldots,X_0) = (1+r^2)E(X_1^2|\ldots,X_0) - r^2X_0^2 \tag{16}$$

We now substitute expression (16) in (6) as follows. Taking the conditional expectation $E(\cdot|\ldots,X_0)$ of both sides of (6), with $k=1$ and substituting (3), we get

$$E(X_1^2|\ldots,X_0) = AX_0^2 + AE(X_2^2|\ldots,X_0) + BX_0^2r^2 + C + D(1+r^2)X_0$$

Replacing $E(X_2^2|\ldots,X_0)$ by the right hand side of (16) we get (14). $\square$

The following result serves as a lemma but is of independent interest.

Proposition 4.1 *Suppose (X_k) is a square-integrable standardized stationary homogeneous Markov chain such that the correlation coefficients satisfy $r \neq 0, 2|r| < 1 + r_2$.*

If (X_k) satisfies condition (5) and the conditional variance $Var(X_k|X_{k-1})$ is a quadratic function of X_{k-1} then one of the following condition holds true:

$$Var(X_k|X_{k-1}) = const \tag{17}$$

or

$$Var(X_k|X_{k-1}) = (1-r^2)X_{k-1}^2 \tag{18}$$

Remark 4.1 *Condition (18) implies that $|X_k| = |X_{k-1}|$ for all k, even in the non-Markov case.*

Remark 4.2 *If linear regression condition (5) is weakened to a symmetric pair of conditions $E(X_k|X_{k-1}) = rX_{k-1}$ and $E(X_{k-1}|X_k) = rX_k$ then the conditional variance can be given by other quadratic expressions, see Example 5.1.*

Proof of Proposition 4.1. If $Var(X_k|X_{k-1})$ is quadratic then there are constants a, b, c such that

$$E(X_k^2|X_{k-1}) = aX_{k-1}^2 + bX_{k-1} + c \tag{19}$$

Since (X_k) is a homogeneous Markov chain and (12) holds true

$$E(X_{k+1}^2|X_{k-1}) = E(aX_k^2 + bX_k + c|X_{k-1}) = a^2X_{k-1}^2 + (a+r)bX_{k-1} + (a+1)c \tag{20}$$

On the other hand, condition (5) implies, see (16)

$$(1+r^2)E(X_k^2|X_{k-1}) = r^2X_{k-1}^2 + E(X_{k+1}^2|X_{k-1}) \tag{21}$$

Combining this with (19) and (20) we get

$$(1+r^2)aX_{k-1}^2 + (1+r^2)bX_{k-1} + (1+r^2)c = (a^2+r^2)X_{k-1}^2 + (a+r)bX_{k-1} + (a+1)c \tag{22}$$

Since $E(X_{k-1}) = 0$ and $E(X_{k-1}^2) = 1$ therefore X_{k-1} must have at least two values. We consider separately two cases.

(a) If X_k has only two values then by Proposition 3.1 $X_k = \pm 1$ and $Var(X_k|X_{k-1}) = 1 - r^2$ is a non-random constant, ending the proof.

(b) If X_{k-1} has at least three values, then $X_{k-1}^2, X_{k-1}, 1$ are linearly independent. Therefore (22) implies

$$(1+r^2)a = a^2 + r^2, (1+r^2)b = (a+r)b, (1+r^2)c = (a+1)c \tag{23}$$

Since (19) implies that $a + c = 1$, the only solutions of (23) are $c \neq 0, a = r^2$ or $c = 0, a = 1$. Since $0 < |r| < 1$, both solutions imply $b = 0$.

Clearly, $a = r^2$ implies (17). On the other hand if $c = 0$ and $a = 1$, then $E(X_k^2|X_{k-1}) = X_{k-1}^2$. Thus (18) hold true.

$\square$

Proof of Theorem 2.1. We first consider the two-valued case. If X_{k-1}^2 is a non-random constant, then $X_{k-1}^2 = 1$ and thus Q is non-unique; one can take $Q(x,y) = (x^2+y^2)/2$ to satisfy (8), or one can take $Q(x,y) = \frac{r^2}{1+r^4}(x^2+y^2) + \frac{(1-r^2)^2}{1+r^4}$ to satisfy (7).

Suppose now that X_k has more than two values. We first verify that that the collusion (8) holds true when $A = 1/(1+r^2)$. In this case the left hand side of (14) is zero. Since X_k has more than two values, this implies that $D = 0$ and $C = 0$. Therefore (4) implies (8).

Now consider the case when $A \neq 1/(1+r^2)$. From (14) we have

$$E(X_k^2|X_{k-1}) = \frac{A(1-r^2)+Br^2}{1-A(1+r^2)} X_{k-1}^2 + \alpha X_{k-1} + \beta \tag{24}$$

where $\alpha = \frac{D(1+r)}{1-A(1+r^2)}$. This shows that $Var(X_k|X_{k-1})$ is quadratic. By Proposition 4.1 we have $\alpha = 0$; since $|r| < 1$ this implies that $D = 0$. We also know that either (17) holds true, which is equivalent to $E(X_k^2|X_{k-1}) = r^2 X_{k-1}^2 + 1 - r^2$, or (18) holds true, which is equivalent to $E(X_k^2|X_{k-1}) = X_{k-1}^2$. We now compare these two expressions with (24): since $\alpha = 0$ and X_{k-1}^2 is non-constant, the coefficients at X_{k-1}^2 must match. That is, either $\frac{A(1-r^2)+Br^2}{1-A(1+r^2)} = r^2$ or $\frac{A(1-r^2)+Br^2}{1-A(1+r^2)} = 1$. By a simple algebra the former implies (7) and the latter implies (8). $\square$

Lemma 4.4 *Suppose that $E(X) = E(Y) = 0, E(X^2) = E(Y^2) = 1, E(X^4) = E(Y^4) < \infty$ and the following conditions hold true*

- $E(Y|X) = rX$
- $E(X^3|Y) = \alpha Y^3 + \beta Y$
- $\alpha \neq r$

Then $\frac{\beta}{r-\alpha} \geq 1$.

Proof. Conditioning in two different directions in EX^3Y we get $rEX^4 = \alpha E(Y^4) + \beta E(Y^2)$. Therefore $E(X^4) = \frac{\beta}{r-\alpha}$. Since $E(X^4) \geq (E(X^2))^2 = 1$ we have $\frac{\beta}{r-\alpha} \geq 1$, which ends the proof. $\square$

The following lemma is based on estimates from Bryc(1995), Theorem 6.2.2. The proof is omitted.

Lemma 4.5 *Suppose X, Y are square-integrable random variables with the same distribution. Let $r = corr(X,Y)$ denote the correlation coefficient and assume that $r \neq 0, \pm 1, E(X|Y) = rY, E(Y|X) = rX, Var(X|Y) = 1-r^2, Var(Y|X) = 1-r^2$. Then $E(X^4) \leq 32\frac{r^2+2|r|+2}{(1-|r|)r^4}$.*

Proof of Theorem 2.2. Since the conclusion is trivially true when $A = 1/(1+r^2)$, throughout the proof we assume that $A \neq 1/(1+r^2)$. In this case (14) implies $Var(X_k|X_{k-1}) = 1-r^2$. Since the assumptions are symmetric, and $0 < |r| < 1$ by Lemma 4.5 and stationarity we have $E(X_1^4) = E(X_2^4) < \infty$.

Notice that (14) implies $E(X_2^2|X_0) = E^{X_0}(E(X_2^2|\ldots,X_1) = r^2E(X_1^2|X_0) + 1 - r^2$. Thus

$$E(X_2^2|X_0) = r^4X_0^2 + 1 - r^4 \tag{25}$$

We now compute conditional moments using the approach of Plucinska(1983). Using constant conditional variance and (1), we write $E(X_1X_2^2|X_0)$ in two different ways as

$$E(E(X_1X_2^2|\ldots,X_0,X_1)|X_0) = E(r^2X_1^3 + (1-r^2)X_1|X_0)$$

and as

$$E(E(X_1X_2^2|X_2,X_0)|X_0) = \frac{r}{1+r^2}E(X_2^2(X_2+X_0)|X_0)$$

Combining these two representations and using (25), and $r \neq 0$ we get after simple algebra

$$rE(X_1^3|X_0) = \frac{1}{1+r^2}E(X_2^3|X_0) + \frac{r^4}{1+r^2}X_0^3 \tag{26}$$

Similarly, we rewrite $E(X_1^2X_2|X_0)$ in two different ways as

$$E(E(X_1^2X_2|\ldots,X_0,X_1)|X_0) = rE(X_1^3|X_0)$$

and, using (2), as

$$E(E(X_1^2X_2|X_2,X_0)|X_0) = E((A(X_2^2+X_0^2)+BX_0X_2+C)X_2|X_0)$$

Using (25), after some algebra we get

$$rE(X_1^3|X_0) = r^2(A+Br^2)X_0^3 + AE(X_2^3|X_0) + (B(1-r^4)+Cr^2)X_0 \tag{27}$$

Solving the system of equations (26), (27) for $E(X_1^3|X_0)$ we get

$$E(X_1^3|X_0) = r\frac{A(1-r^2)+Br^2}{1-A(1+r^2)}X_0^3 + \frac{B(1-r^4)+Cr^2}{r(1-A(1+r^2))}X_0 \tag{28}$$

Substituting (4), (7), and denoting $\tilde{A} = A(1+r^2)$ we have

$$E(X_1^3|X_0) = r^3X_0^3 - \frac{1-r^2}{r^3}\frac{\tilde{A}(1+2r^4)-r^2(1+2r^2)}{1-\tilde{A}}X_0 \tag{29}$$

Therefore by Lemma 4.4 and a simple calculation we have

$$\frac{\tilde{A}(1+r^4)-r^2(1+r^2)}{r^4(1-\tilde{A})} \leq 0 \tag{30}$$

Since $\frac{r^2}{1+r^4} < \frac{1}{1+r^2}$ this implies that either $A > 1/(1+r^2)$ or $A \leq \frac{r^2}{1+r^4}$. $\square$

Proof of Theorem 2.3. For $A \neq 1/(1+r^2)$ let

$$q = \frac{r^2 - A(1+r^2)}{r^4(1-A(1+r^2))} \tag{31}$$

The range of values of A implies that $-1 \leq q \leq 1$. We give the proof for the case $-1 < q \leq 1$. The only change needed for the case $q = -1$, is to use the symmetric two-valued Markov chain defined in Section 3 instead of the Markov chain M_k defined below.

Define orthogonal polynomials $Q_n(x)$ by the recurrence

$$Q_{n+1}(x) = xQ_n(x) - (1+q+\ldots+q^{n-1})Q_{n-1}(X) \tag{32}$$

with $Q_0(x) = 1$, $Q_1(x) = x$. Let $\mu(dx)$ denote the probability measure which orthogonalizes Q_n (see eg Chihara(1978), Theorem 6.4), and for fixed $-1 < r < 1$ define

$$P(x,dy) = \sum_{n=0}^{\infty} r^n\tilde{Q}_n(x)\tilde{Q}_n(y)\mu(dy) \tag{33}$$

where $\tilde{Q}_n(x) = Q_n(x)/\|Q_n\|_{L_2(\mu)}$ are normalized orthogonal polynomials Q_n. By Bryc(1998), Lemma 8.1, for $-1 < q \leq 1$ formula (33) defines a Markov transition function with invariant measure μ. For $-1 < q \leq 1$, let M_k be a stationary Markov chain with the initial distribution μ and transition probability $P(x,dy)$.

It is known that μ is either gaussian or of bounded support, see Koekoek-Swarttouw(1994), and hence the joint distribution of $M_1, \ldots, M_d$ is uniquely determined by mixed moments $E(M_1^{k_1} \ldots, M_d^{k_d})$. We will show by induction with respect to d that

$$E(X_1^{k_1} \ldots, X_d^{k_d}) = E(M_1^{k_1} \ldots M_d^{k_d}) \tag{34}$$

for all $d \geq 1$ and all non-negative integers $k_1, ..., k_d$.

By Bryc(1998) marginal distributions are equal, $X_1 \cong M_1$; this shows that equality (34) holds true for all integer $k_1 \geq 0$ when $d = 1$. Suppose (34) holds for all $k_1, ..., k_d \geq 0$. Fix integer $k = k_{d+1} \geq 0$. Expand polynomial x^k into orthogonal expansion, $x^k = \sum_{j=0}^{k} a_j Q_j(x)$. Then

$$E(X_1^{k_1} \ldots, X_d^{k_d} X_{d+1}^k) = \sum_{j=0}^{k} a_j E(X_1^{k_1} \ldots, X_d^{k_d} E(Q_j(X_{d+1})|X_1, \ldots, X_d)$$

Repeating the reasoning that lead to Bryc(1998), Lemma 6.3, we have $E(Q_j(X_{d+1})|X_1, \ldots, X_d) = r^j Q_j(X_d)$. Therefore $E(X_1^{k_1} \ldots X_d^{k_d} X_{d+1}^k) = \sum r^j a_j E(X_1^{k_1} \ldots X_d^{k_d} Q_j(X_d))$ is expressed as a linear combination of moments that involve only $E(X_1^{j_1} \ldots X_d^{j_d})$. Since the same reasoning applies to M_k, we have $E(M_1^{k_1} \ldots M_d^{k_d} M_{d+1}^k) = \sum r^j a_j E(M_1^{k_1} \ldots M_d^{k_d} Q_j(M_d))$, and (34) follows. $\square$

5 Example

This section contains an example of a stationary reversible Markov chain with linear regressions and quadratic conditional moments, which does not satisfy condition (5). The Markov chain has polynomial regressions of all orders, and does not satisfy the conclusion of Proposition 4.1.

Example 5.1 *Suppose $T_n(x)$ are Chebyshev polynomials of the first kind, $T_0 = 1, T_1(x) = x, T_2(x) = 2x^2 - 1, xT_n(x) = \frac{1}{2}T_{n+1}(x) + \frac{1}{2}T_{n-1}(x)$. Let $\mu(dx) = \frac{1}{\pi}\frac{1}{\sqrt{1-x^2}}dx$. Then T_n are orthogonal in $L_2(d\mu)$ and $\|T_0\|^2_{L_2(d\mu)} = 1$ and for $k > 0$ $\|T_k\|^2_{L_2(d\mu)^2} = \frac{1}{2}$. We define transition density by $p(x, y) = \sum_{n=0}^{\infty} r^n T_n(x)T_n(y)$.*

Since $T_n(x) = \cos(n \arccos(x))$, the series can be summed. Writing $T_n(x) = \cos(n\theta_x)$ we have $T_n(x)T_n(y) = \frac{1}{2}\cos(n(\theta_x + \theta_y)) + \frac{1}{2}\cos(n(\theta_x - \theta_y))$ Therefore

$$p(x, y) = \frac{1}{2}\frac{1 - r\cos(\theta_x + \theta_y)}{1 + r^2 - 2r\cos(\theta_x + \theta_y)} + \frac{1}{2}\frac{1 - r\cos(\theta_x - \theta_y)}{1 + r^2 - 2r\cos(\theta_x - \theta_y)}$$

This shows that $p(x, y) \geq \frac{1-|r|}{(1+|r|)^2} > 0$. The expression simplifies to

$$p(x, y) = \frac{1 - r^2 + r(2r(y^2 + y^2) - (3 + r^2)yx)}{(1 - r^2)^2 + 4r^2(x^2 + y^2 - (r + 1/r)xy)}$$

Thus we can define the Markov chain X_k with one-step transition probabilities $P_x(dy) = p(x, y)\mu(dy)$ and initial distribution μ. Since $\int p(x, y)\mu(dx) = 1$, the chain is stationary.

Notice that by the definition of $p(x, y)$ we have $E(T_n(X_1)|X_0) = r^n\|T_n\|_2^2 T_n(X_0)$. Therefore for $n \geq 1$ we have $E(T_n(X_1)|X_0) = \frac{1}{2}r^n T_n(X_0)$

In particular $E(X_1|X_0) = r/2X_0$, and $E(2X_1^2 - 1|X_0) = \frac{1}{2}r^2(2X_0^2 - 1)$. The latter implies $E(X_1^2|X_0) = \frac{1}{2}r^2X_0^2 + \frac{1}{2} - \frac{1}{4}r^2$ and hence the conditional variance $Var(X_1|X_0) = \frac{1}{4}r^2X_0^2 + \frac{1}{2} - \frac{1}{4}r^2$ is non-constant. This should be contrasted with the conclusion of Proposition 4.1 and assumptions in Bryc(1998), Wesolowski(1993).

Acknowledgements I would like to thank Dr. Sivaganesan and W. Matysiak for helpful discussions.

References

[1] Bryc, W., 1995. *Normal distribution: characterizations with applications*, Lecture Notes in Statistics, vol. 100, Springer.

[2] Bryc, W., 1998. *Stationary random fields with linear regressions*, preprint WWW: `http://ucms02.csm.uc.edu/preprint`

[3] Bryc, W., & Plucińska, A. A characterization of infinite gaussian sequences by conditional moments. Sankhya A vol. 47 (1985), 166–173.

[4] Chihara, T. S., 1978. *An introduction to orthogonal polynomials*, Gordon and Breach, New York.

[5] Koekoek, R. & Swarttouw, R. F., 1994. The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue, Report 94-05 Technische Universiteit Delft, WWW: `http://www.can.nl/~demo/AWscheme/intro.html`

[6] Plucińska, A. *On a stochastic process determined by the conditional expectation and the conditional variance*. Stochastics 10(1983), 115–129.

[7] Szabłowski, P. Can the first two conditional moments identify a mean square differentiable process? Computers Math. Applic. 18 (1989), 329–348.

[8] Wesołowski, J. *Characterizations of some Processes by properties of Conditional Moments*. Demonstratio Math. 22 (1989), 537–556.

[9] Wesołowski, J. Stochastic Processes with linear conditional expectation and quadratic conditional variance. Probab. Math. Statist. (Wrocław) 14 (1993), 33–44.